EXHIBIT A

0273-2289/99/77–79/0055/$12.75

Two High-Throughput Techniques for Determining Wood Properties as Part of a Molecular Genetics Analysis of Hybrid Poplar and Loblolly Pine

GERALD TUSKAN,*[,1] DARRELL WEST,[1] HARVEY D. BRADSHAW,[2] DAVID NEALE,[3] MITCH SEWELL,[3] NICK WHEELER,[4] BOB MEGRAW,[4] KEITH JECH,[4] ART WISELOGEL,[5] ROBERT EVANS,[5] CAROLYN ELAM,[5] MARK DAVIS,[5] AND RON DINUS[6]

[1]P.O. Box 2008, MS-6422, Oak Ridge National Laboratory, Oak Ridge, TN 37831, E-mail: gtk@ornl.gov; [2]University of Washington, Seattle, WA; [3]U. S. Forest Service, Davis, CA; [4]Weyerhaeuser Company, Centralia, WA; [5]National Renewable Energy Laboratory, Golden, CO; and [6]University of British Columbia, BC, Canada

Abstract

Two new high-throughput techniques, computer tomography X-ray densitometry (CT scan) and pyrolysis molecular beam mass spectrometry (pyMBMS), coupled with quantitative trait loci (QTL) analysis, were tested as a means to overcome the time and cost associated with conventional characterization of biomass feedstock components. Applications of these two techniques were evaluated using hybrid poplar for the CT scan and loblolly pine for the pyMBMS. Segregating progeny from hybrid poplar varied in specific gravity, with individual mean estimates ranging from 0.21–0.41. Progeny from loblolly pine varied in lignin, α cellulose, and mannan contents, with individual mean estimates of lignin content ranging from 28.7–33.1%, α cellulose content from 28.8–43.5% and mannan content from 4.2–10.1%. QTL analysis of the loblolly pine data suggested that eleven QTLs were associated with individual feedstock characteristics and that two QTLs for several feedstock components were linked to the same position on the loblolly pine genetic map. Each QTL individually accounted for 7–13% of the total phenotypic variation in associated loblolly pine feedstock components.

Index Entries: Pyrolysis molecular beam mass spectrometry; computer tomography; genetic markers; QTL; lignin; cellulose; hemicellulose.

*Author to whom all correspondence and reprint requests should be addressed.

Introduction

Feedstock composition affects product value, product quality, and production efficiency in all biomass-based industries. For example, in the pulp and paper industry, pulp yields are known to increase with increased cellulose content, pulp cohesiveness is positively associated with hemicellulose content, and energy consumption during chemi-thermomechanical pulping is inversely related to lignin content *(1)*. In biochemical conversion of lignocellulosic feedstock to ethanol fermentation, efficiency is directly related to cellulose:hemicellulose ratios, biocatalyst activity can be inhibited by feedstock extractives, and the cost and types of feedstock pretreatment are strongly influenced by lignin content and composition *(2)*. In the biomass power industry, slagging of boilers during direct combustion of biomass feedstocks is positively associated with feedstock ash content *(3)*.

Despite the importance of feedstock composition on each of the aforementioned processes or products, very little has been done genetically to improve feedstock quality in the biomass-based industries *(4)*. Two reasons account for this lack of selection and improvement. First, conventional analytical methods often require months of process time per set of samples, with each analysis costing hundreds of dollars. Estimates of total composition are often inaccurate because of incompatibility of assays, degradation of component fractions during specific assays, or the need to use subtractive methods to calculate certain feedstock fractions. Secondly, genetic selection for improved feedstock quality has been hampered by inaccurate phenotypic data, lengthy assays, and long generation intervals for most perennial species. Because genetic selection generally relies on phenotypic characterization of hundreds of progeny replicated over time and space, the cost and the delays associated with estimates of feedstock quality from conventional wet chemistry prohibit such assessments as part of traditional breeding programs.

Two new high-throughput analytical approaches, coupled with genetic mapping and marker-assisted selection methods *(5)*, promise a means to overcome the aforementioned constraints. The two techniques are computer tomography X-ray densitometry (CT scan) and pyrolysis molecular beam mass spectrometry (pyMBMS). The objectives of this article are 1. to describe the general application of CT scan to segregating hybrid poplar progeny and pyMBMS to segregating loblolly pine progeny and 2. to relate the results for loblolly pine to the existing genetic map through a quantitative trait loci (QTL) analysis.

PyMBMS combined with multivariate analysis has been used previously to estimate cell-wall chemical composition of different biological materials *(6–8)*. These studies have applied pyrolysis mass spectrometry techniques to a variety of lignocellulosic materials including wood, herbaceous fiber, and pulps, and have found a correlation between the pyrolysis results and the amount of cell-wall constituents determined by conventional techniques. Agblevor et al. *(6)* have demonstrated that the combina-

tion of pyMBMS and factor analysis can reliably estimate plant cell-wall components. Agblevor et al. *(6)* observed good correlation between total pentosans or total hexosans obtained using conventional wet chemistry methods and pyMBMS factor scores. However, the pyMBMS/lignin correlation was very weak, presumably because of degradation products that interfered with the Klason lignin analysis. For this study, we propose to use projection to latent structure analysis to regress conventional wet chemistry analysis with pyMBMS spectrometric data *(9)*. The advantage of the projection of latent structure data reduction approach is that the analysis can handle several responses (i.e., chemical/feedstock components) simultaneously, rather than regressing a single component with a single spectral peak or group of peaks.

Materials and Methods

Plant Materials

A three-generation inbred hybrid poplar pedigree containing 375 segregating progeny was grown through age 4 by Boise Cascade Corp. in eastern Oregon. *Populus trichocarpa* was used as the maternal grandparent and *Populus deltoides* as the paternal grandparent. The segregating progeny were obtained by crossing a single full-sibling male with a single full-sibling female progeny from the initial parental mating. A single 5-mm increment core representing both sides of a single tree, passing through or near the pith, was extracted 1.5 m above ground level from each member of the pedigree, including the parents and grandparents, and were shipped to Oak Ridge National Laboratory (Oak Ridge, TN) for further processing. Similarly, a three-generation, outcrossed loblolly pine (*Pinus taeda* L.) pedigree containing 172 segregating progeny was field-grown through ages 11–13 by Weyerhaeuser Company in North Carolina and Oklahoma. Loblolly pine increment cores were sampled as described for hybrid poplar and were shipped to Centralia, WA for further processing. In addition, large sections of stem segments were removed from individuals known to have extreme specific gravity phenotypes *(10)* and used as a source of materials for conventional compositional analyses. Increment cores for both species were dried under vacuum at room temperature and stored at 0°C until needed for CT scan or pyMBMS.

CT Scan: Hybrid Poplar

Members of the *Populus* genus are represented by diffused porous species with moderate to low specific gravity and a lack of clearly defined annual growth rings *(11)*. The CT scans were used to determine both annual ring boundaries and within-ring specific gravity. Each increment core was removed from storage, dried under vacuum at room temperature, and positioned on a wooden mounting block for X-ray scanning. For CT scan, each core was subjected to 420 kV, 3 mA collimated, 0.474-mm thick X-ray

beam using a Scientific Measurement Systems Model B201 CAT scanning device *(12)*. Each mounting block containing approximately 30 cores was translated and rotated 10°/min over an 18-min period in the X-ray beam in order to obtain a set of density functions. The density at any point within 100 μm in the 0.474 mm beam was determined by simultaneously solving the set of density functions. The X-ray density data from the reconstructed three-dimensional digital image for each core was stored as an array of numbers representing local X-ray attenuation that was post-scan processed and converted to specific gravity by an algorithm developed from volumetric measurements and direct radiographic methods. Reconstructed specific gravity images from each core were used to delineate annual growth rings.

PyMBMS: Loblolly Pine

The fifth-yr growth ring from each loblolly pine increment core was removed, ground in a Wiley mill, and screened through a 0.2-mm mesh. These ground samples were then shipped at room temperature, next-day delivery to the National Renewable Energy Laboratory (Golden, CO) for pyMBMS analysis. Three 15 mg subsamples from each sampled genotype (i.e., from all progeny, parents, and grandparents) were removed from storage and used in the analysis. Each subsample was placed in a quartz boat and introduced into a pyrolysis reactor at 550°C under a helium flow rate of 5.0 L/min at 21°C, 0.1 MPa. Vacuum expansion created a molecular beam that was introduced into an Extrel Model TQMS C50 mass spectrometer (electron impact ionization voltage = 22.5 eV), which measured atomic mass between 15 and 300 Da in real time during the pyrolysis period *(13)*. Total run time/subsample was 1.5 min; all subsamples were analyzed over a consecutive period of time to avoid temporal variability in analytical equipment. Spectral data from the mass spectrometer were then analyzed using multivariate statistics.

Genotypes of known extreme phenotypes were selected and used as external calibration standards during the pyMBMS analysis. That is, information on an extractive-free-basis for lignin, α cellulose, and hemicellulose components, percent extractives, and ash contents were estimated independently by laboratories at PAPRICAN (Pointe-Claire, Quebec, Canada) and Michigan Tech University (Houghton, MI) for each set of external calibration standards. Lignin content was determined by the Klason method, TAPPI T222 *(14)*. Additional loblolly pine calibration standards were provided by Weyerhaeuser Company (Tacoma, WA). Results from each laboratory were compared and inconsistencies resolved before initiating the pyMBMS analysis. Factor scores from the correlation matrix on normalized mass spectral data for each external calibration standard were correlated with known feedstock composition. Projection of latent structure models, in the form $Y = \beta_0 + \beta_1$ (principal component X), were developed for each feedstock component and used to predict feedstock composition for each tested genotype.

QTL Analysis: Loblolly Pine

QTL were characterized using a simple regression method that utilizes the three-generation structure of the outbred loblolly pine pedigree *(15,16)*. This analysis simultaneously uses genetic information from all markers of a linkage group to provide a robust test for the presence of genetic variation associated with each feedstock component. At regular genomic intervals, the probability of each progeny being one of the four possible genotypic classes, determined as a function of the grand parental genotypes, was calculated. The phenotypes were then regressed onto these genotypic probabilities using a multiple regression approach *(17)*. Linear regression is relatively simple compared with a maximum likelihood approach and therefore allows for testing more complex (and potentially more realistic) genetic models (e.g., fitting one vs two QTLs/linkage group). On a genome-wide basis, QTLs were identified at either a "suggestive" at $p \leq 0.01$ *(18)* or significant at $p \leq 0.005$ level. QTLs were then positioned onto the loblolly pine consensus map *(19)*. This consensus map was constructed from two independent three-generation pedigrees and contains 357 restriction fragment-length polymorphism, random amplified polymorphic DNA, and isozyme genetic markers. The linkage information from these two pedigrees was coalesced into 12 integrated linkage groups (and several smaller groups) representing approximately 85% of the loblolly pine genome.

Results and Discussion

Wood Property Phenotypes

Wood-Specific Gravity

Estimates of wood-specific gravity based on CT scan varied widely within samples, across annual growth rings (Fig. 1), and among all sampled genotypes (Fig. 2). Increment cores extracted at 1.5 m typically contained three discrete growth rings, with wood-specific gravity increasing across the growing season within each ring (Fig. 1). The highest wood-specific gravity/yr occurred generally at the end of the growing season, whereas the lowest occurred at the beginning of each growing season. There was no overall trend among years, such that average yr 1 specific gravity was not consistently higher or lower than yr 2 specific gravity from one increment core to the next. Alternate sides of a single increment core displayed subtle differences in specific gravity, though generally maximum and minimum values were consistent within rings. However, specific gravity values occasionally fluctuated unexpectedly within annual growth rings. This fluctuation, visually observed as layers of cells with increased cell-wall thickness, may have been caused by changes in water or nutritional status of each tree or possibly owing to insect or disease infestations known to occur each growing season.

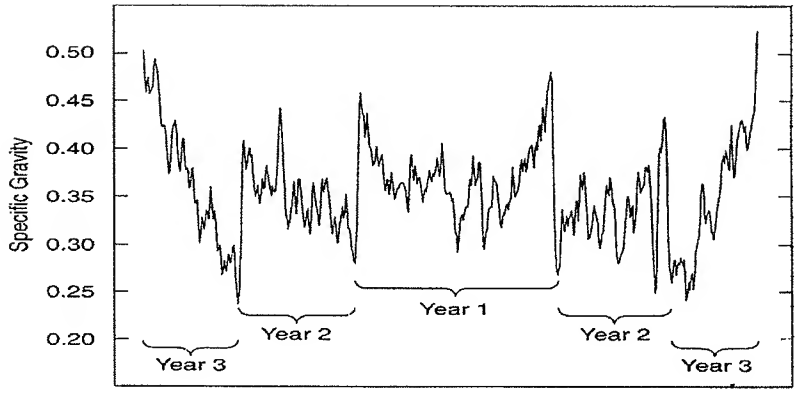

Fig. 1. Computer tomography X-ray densitometry results from a single increment core extracted at 1.5 m from a 4-yr-old hybrid poplar tree grown in eastern Oregon under fertigation. The increment core and CT scan represent both sides of a single tree starting with the pith in the middle of the image.

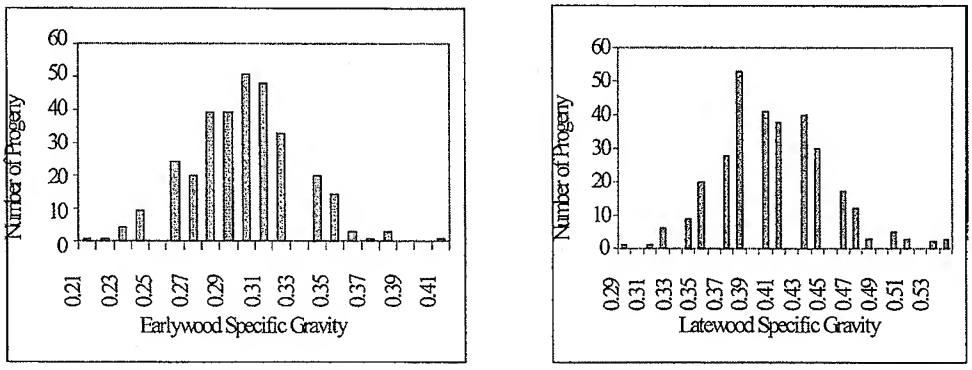

Fig. 2. Frequency of earlywood and latewood specific-gravity estimates based on computer tomography X-ray densitometry for the penultimate growth ring from 375 4-yr-old hybrid poplar progeny grown in eastern Oregon.

Mean "earlywood" specific gravity for the hybrid poplar pedigree was 0.29, ranging from 0.21–0.41 (Fig. 2). Note that true earlywood does not occur in a diffuse-porous species such as hybrid poplar. We referenced the mean increase in density that occurs throughout the growing season (as previously noted) as a point that differentiates earlywood from "latewood." Mean latewood specific gravity was 0.40, ranging from 0.29–0.54. The phenotypic correlation between these two estimates was 0.77, suggesting that genotypes that have high earlywood specific gravity will likely have high latewood specific gravity. For the *P. trichocarpa* grandparent, estimates of earlywood and latewood specific gravity averaged 0.24 and 0.34, respectively, vs 0.27 and 0.36, respectively, for the *P. deltoides* grandparent. The parental-generation specific gravity estimates were consistently

Table 1
Feedstock Composition as Determined
by Conventional Analytical Chemistry for 6 of the 26 Loblolly Pine Samples
Used as Calibrated External Standards During pyMBMS Analysis

		% Extractive-free oven dry weight			
Sample ID	Wood type	Lignin	α Cellulose	Hemicellulose	Ash
97-1	Juvenile	34.7	35.8	25.5	0.34
	Mature	32.3	36.2	24.4	0.30
97-2	Juvenile	32.1	35.7	25.8	0.32
	Mature	32.0	36.1	25.9	0.30
97-3	Juvenile	31.2	37.4	25.5	0.31
	Mature	30.2	38.9	25.7	0.28

higher than the grandparental generation estimates. All estimated specific gravity values (Fig. 2) were in the range typically reported for *Populus* species *(3,11,20)*.

Wood Composition

Estimates of feedstock composition by conventional analytical approaches for the loblolly pine used as external calibration standards ranged from 26.7–34.7% for lignin, 35.7–48.0% for α cellulose, 24.4–25.9% for hemicellulose, and 0.28–0.34% for ash content (Table 1). Mature wood/genotype tended to have lower lignin content, higher α cellulose content, and lower ash content as compared with juvenile wood from the same sample. Individual genotypes also varied in extreme values, with sample 97-1 providing the highest estimate of lignin and sample 97-3 providing the highest cellulose values. Sample 97-2 tended to have average values for all tested components. Estimates for each component were determined separately without resorting to corrective subtraction, and thus, component values do not sum to 100. Based on these results, estimates for individual feedstock components were then used to calibrate the pyMBMS factor scores.

PyMBMS projection of latent structure models were used to estimate lignin, α cellulose, glucan, specific gravity, extractives, and some individual hemicellulose sugar components, xylans, galactans, and mannans. The range of the chemical composition for each component, the correlation coefficient, and the root mean squared error of prediction (RMSEP) of the projection of latent structure models used to predict the loblolly pine chemical compositions are shown in Table 2. The arabinan concentrations could not be estimated using pyMBMS methods owing to its low concentration and difficulty obtaining accurate arabinose estimates using conventional wet chemistry methods. The predictive model was derived from 26 samples run in triplicate. The RMSEP is the average uncertainty that can be expected when predicting the concentrations for the unknown loblolly pine samples using this model. The results of the future predictions can be expected to be

Table 2
Range of Phenotypic Standards, Correlation Coefficients, and Root Mean Squared Errors of Prediction (RMSEP) of the Loblolly Pine pyMBMS Projection of Latent Structure Regression Model Used to Predict Feedstock Composition

Chemical component	Range	Correlation coefficient	RMSEP
Lignin	26.7–29.3	0.92	0.34
α Cellulose	41.5–48.0	0.95	0.89
Extractives	1.0–2.9	0.91	0.21
Specific gravity	0.38–0.57	0.96	0.017
Glucans	46.5–55.5	0.85	1.3
Galactan	2.6–4.0	0.83	0.12
Mannan	8.8–11.9	0.88	0.47
Xylan	7.0–8.3	0.86	0.20

Table 3
Summary of pyMBMS Estimates of the Chemical Composition for a Segregating Pedigree of Loblolly Pine Sampled at the Fifth-Yr Growth Ring

Chemical component	Phenotypic mean	Range	SD
Lignin	30.8	28.7–33.1	0.8
α Cellulose	36.4	28.8–43.5	2.7
Extractives	1.6	0.8–2.1	0.2
Specific gravity	0.27	0.10–0.43	0.06
Galactan	2.2	1.1–3.0	0.3
Mannan	7.5	4.2–10.1	1.0
Xylan	9.0	8.0–10.1	0.4

within two RMSEP. The pyMBMS predictive model was reconciled using a full cross-validation scheme.

Based on results of the projection of latent structure models, estimates of the overall chemical composition of the loblolly pine fifth-year growth ring for the different wood components are shown in Table 3. Figure 3 shows the average triplicate estimates for α cellulose, lignin, and a hemicellulose component, mannan, determined using the pyMBMS analysis. The pyMBMS analysis clearly distinguishes between the wood sampled from the earlywood and latewood portion of the 5th-yr growth ring. The pyMBMS analysis correctly indicates a higher lignin content (Fig. 3B) and lower specific gravity (data not shown) in the earlywood portion of the ring with respect to the latewood portion of the ring consistent with previous reports *(21,22)*. The analysis also determined that the concentrations of mannan (Fig. 3C) and galactose (data not shown) are lower in the earlywood than in the latewood, again, consistent with previous studies of softwoods *(23)*. The data also defines differences in the average chemical composition of the samples collected from trees grown in North Carolina

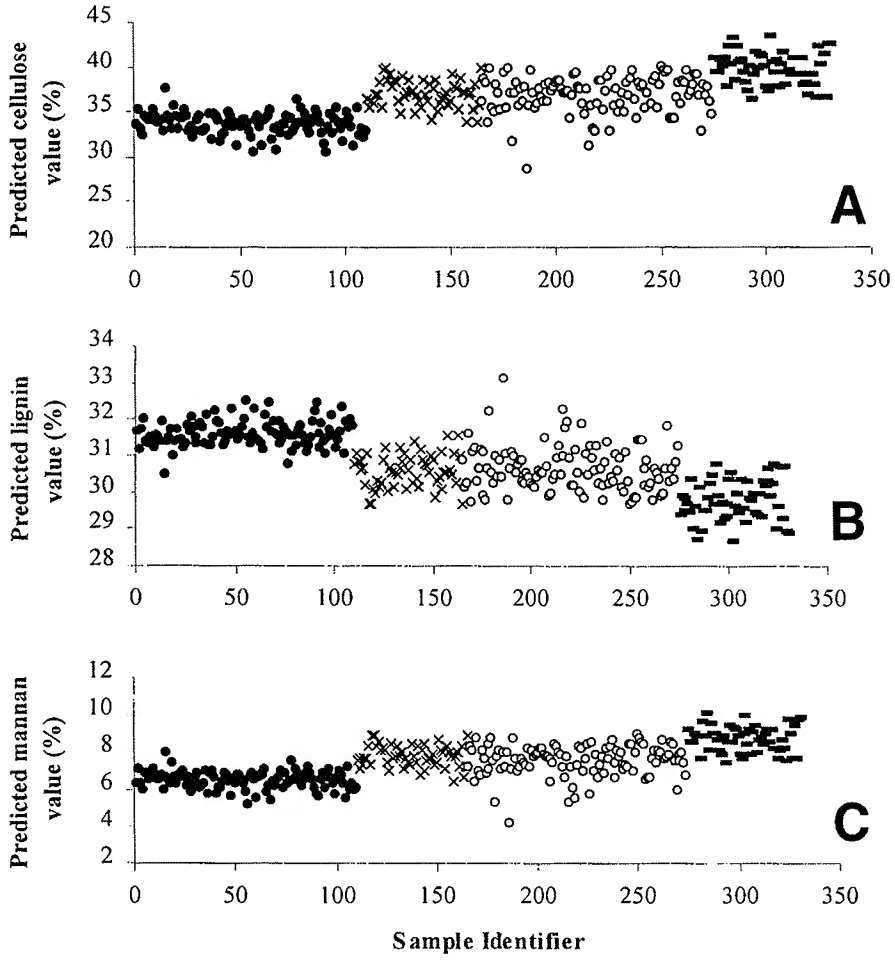

Fig. 3. PyMBMS predicted values for **(A)** α cellulose, **(B)** lignin, and **(C)** mannan. The earlywood and latewood samples from the North Carolina site are represented by ● and ○, respectively. The earlywood and latewood samples from the Oklahoma site are represented by x and –, respectively.

from those grown in Oklahoma (Fig. 3). The reason for these differences is not apparent. Finally, pyMBMS analysis showed that the grand parental and parental generations provided estimates of lignin content clustered within one-half standard deviation around the overall progeny mean (data not shown). Similar results were found for all other compositional components.

QTL Analysis

QTL were identified that were associated with each feedstock component examined in loblolly pine. From fifth-year growth ring pyMBMS data, two QTLs were identified for both earlywood and latewood lignin content, one mapping to the same position for both traits. These QTLs explain roughly 10% of the total variation for both lignin traits. Two additional

QTLs were associated with glucan content in earlywood, accounting for ~13% of the total variation. A single QTL was associated with both earlywood and latewood hemicellulose content and each accounted for approximately 7% of the variation in hemicellulose. Interestingly, two positions on the loblolly pine genetic map, one on linkage group 8 and the other on linkage group 16, were associated with a large QTL cluster of several feedstock components. The QTLs on linkage group 8 were related to galactans, mannans, xylans, and lignin contents from latewood. The QTLs on linkage group 16 were related to earlywood and latewood galactans, mannans, xylans, and lignin contents. It may be possible that there are genes within the limits of the tested loblolly pine genome that influence overall cell-wall formation and are thus linked to individual cell-wall component estimates. If verified in a second, unrelated loblolly pine pedigree, the aforementioned QTLs explain a sufficiently high enough portion of the total variation in each feedstock component to merit a test of a marker-aided selection scheme for improved wood properties.

In summary, CT X-ray densitometry and pyMBMS have proven to be reliable, accurate, and affordable methods for characterizing feedstock properties. Initial success with CT scan for hybrid poplar and pyMBMS for loblolly pine are currently being evaluated in reciprocal studies, i.e., CT scan for loblolly pine and pyMBMS for hybrid poplar. In the examined loblolly pine and hybrid poplar pedigrees, progeny from both species displayed genetic variation in all examined feedstock components. QTL analysis suggested that there are positions within the genome of loblolly pine that are associated with feedstock composition and that many of the QTLs are tightly linked to one another. As such, these new techniques should facilitate genetic assessment of individual feedstock components, as well as genetic selection aimed at wood property trait improvement.

Acknowledgments

The authors acknowledge the U.S. Department of Energy's Agenda 2020 Program, the Bioenergy Feedstock Development Program, and the Biomass Power program for funding used to support the collection of the phenotypic data, data analysis, and manuscript preparation. The authors also acknowledge Boise Cascade Corp. (Wallula, WA) and Weyerhaeuser Company (Centralia, WA) for allowing us use of hybrid poplar and loblolly pine plant materials, and the Poplar Molecular Genetic Cooperative (Seattle, WA) and the U.S. Forest Service, who provided funds for the construction of the hybrid poplar and loblolly pine genetic maps, respectively. Manuscript editing was provided by Rachel M. Tuskan. The National Renewable Energy Laboratory is managed by Midwest Research Institute for DOE under contract DE-AC36-83CH10093. Oak Ridge National Laboratory is operated for DOE under contract DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp., Environmental Sciences Division, Oak Ridge National Laboratory, Publ. no. 4850.

References

1. Smook, G. A. and Kocurek, M. J. (1988), *Handbook for Pulp and Paper Technologists.* TAPPI, Atlanta, GA.
2. Duff, S. J. B. and Murray, W. D. (1996), *Biores. Technol.* **55,** 1–33.
3. Dinus, R. J., Dimmel, D. R., Feirer, R. P., Johnson, M. A., and Malcolm, E. W. (1990), *ORNL Report, 88-SC006,* U.S. Government Press (available through National Technical Information Service U.S. Department of Commerce), Oak Ridge, TN.
4. Zobel, B. J. and Jett, J. B. (1997), *Genetics of Wood Production,* Springer-Verlag, New York.
5. Williams, C. G. and Neale, D. B. (1992), *Can. J. For. Res.* **22,** 1009–1017.
6. Agblevor, F. A., Evans, R. J., and Johnson, K. D. (1994), *J. Anal. Appl. Pyrolysis* **30,** 125– 144.
7. Windig W., Heckler, C. E., Agblevor, F. A., and Evans, R. J. (1992), *Chemometr. Intell. Lab. Syst.* **14,** 195–207.
8. Windig, W., Meuzelarr H. L. C., Shafizadeh, F., and Kelsey, R. G. (1984), *J. Anal. Appl. Pyrolysis* **6,** 233–250.
9. Martens, H. and Naes, T. (1989), *Multivariate Calibration,* Wiley, New York.
10. Williams, C. G. and Megraw, R. A. (1994), *Can. J. For. Res.* **24,** 714–722.
11. Core, H. A., Cote, W. A., and Day, A. C. (1979), *Wood Structure and Function,* Syracuse University Press, Syracuse, NY, pp. 72–75.
12. Elliott, J. C., Anderson, P., Davis, G. R., and Leng, F. S. (1994), *JOM* **43,** 11–19.
13. Davis, M. F., Johnson, D. K., Agblevor, F., Fennell, J., and Ashley, P. (1995), *Proceedings of the 2nd Biomass Conference of the Americas,* NREL/CP-200-8098, Golden, CO, pp. 216–225.
14. Lapierre, C., Pollet, B., and Rolando, C. (1995), New insights into the molecular architecture of hardwood lignin. *Res. Chem. Intermediates* **21,** 397–412.
15. Haley, C. S. and Knott, S. A. (1992), *Heredity* **69,** 315–324.
16. Knott, S. A. (1997), *TAG* **94,** 810–820.
17. Haley, C. S., Knott, S. A., and Elsen, J. M. (1994), *Genetics* **136,** 1195–1207.
18. Lander, S. and Kruglyak, D. (1995), *Nat. Genet.* **11,** 241–247.
19. Sewell, M., Sherman, B. K., and Neale, D. B. (1999), *Genetics,* in press.
20. Goyal, G. C., Fisher, J. J., Krohon, M. J., Packwood, R. E., and Olson, J. R. (1997), *TAPPI Proceedings,* TAPPI Press, Atlanta, GA, pp. 857–862.
21. Fengal, D. and Wegener, G. (1984), *Wood: Chemistry, Ultrastructure, Reactions,* Walter de Gruyter, Berlin, p. 59.
22. Browning, B. L. (1975), In, *The Chemistry of Wood,* Krieger, R. E., ed., Huntington, NY, p. 67.
23. Sjöström, E. (1981), *Wood Chemistry, Fundamentals and Applications,* Academic, New York, p. 66.